.D-A245 541

DTIC
S ELECTE D
FEB 06 1992
D

This document has been approved for public release and sale; its distribution is unlimited.

92-03010

Best Available Copy

GAO

United States
General Accounting Office
Washington, D.C. 20548

Information Management and
Technology Division

DTIC
ELECTE
FEB 06 1992
S D D

B-240617

April 8, 1991

The Honorable Albert Gore, Jr.
Chairman, Subcommittee on Science,
Technology, and Space
Committee on Commerce, Science,
and Transportation
United States Senate

Dear Mr. Chairman:

Your office requested that we (1) identify current and advanced data storage technologies to help the Committee evaluate their use for the National Aeronautics and Space Administration's (NASA) future storage needs, and (2) determine NASA's initiatives for using these technologies. On September 12, 1990, we provided you with a report addressing the first objective, which identified a number of storage technologies, their general characteristics, and costs.[1] This report responds to the second objective. It also discusses several challenges NASA will face in the 1990s and beyond, the visibility of these important initiatives to the Congress, and examines whether opportunities exist for reducing development costs on the initiatives. Details of our objectives, scope, and methodology are provided in appendix I.

Results in Brief

NASA missions will soon generate quantities of data never before encountered. Between 1990 and 2000, the volume of data archived will rise by over 5,600 percent. Because the information systems now in place will be unable to support NASA's future processing and storage requirements, NASA is pursuing a number of major projects and initiatives that will use advanced data storage technologies. For example, to accommodate much of its future storage requirements, NASA plans to design and construct a complex information system called the Earth Observing System Data Information System (EOSDIS). Designed to support the estimated $38 billion Earth Observing System (EOS) program, EOSDIS will use an advanced mass storage system—not yet developed—to meet its anticipated need for storing and retrieving more than 1,000 times the amount of information now stored in the Library of Congress.

Accesion For
NTIS CRA&I ✓
DTIC TAB
Unannounced
Justification
By
Distribution /
Availability Codes
Dist | Avail and / or Special
A-1

This document has been approved for public release and sale; its distribution is unlimited.

[1] Space Data: Information on Data Storage Technologies (GAO/IMTEC-90-88FS, Sept. 12, 1990).

While initiatives such as EOSDIS could help the agency manage and store the volumes of data from future missions, these initiatives pose formidable challenges. For example, these initiatives may rely on an advanced mass data storage system that is not yet available, may not work, and for which no media and data format standards exist. According to NASA and the National Research Council, developing an information system of the size and complexity of EOSDIS will truly be a pioneering effort.

Given the cost and importance of these initiatives and the challenges they pose, their visibility to the Congress is vital. However, this may not happen because NASA does not currently plan to report separately on each initiative.

Finally, in managing initiatives like EOSDIS, NASA must be vigilant for ways to minimize costs. For example, we found that some of the storage requirements currently planned for EOSDIS may be unnecessarily redundant, adding about $35 million to the estimated cost of the program. Although EOSDIS officials told us they would modify the design documents to eliminate the unnecessary redundancy, they had not done this when we completed our work in December 1990.

Background

Since 1958, NASA has spent billions on its space science program and successfully launched over 260 scientific missions. Data from these missions have expanded our understanding of the earth, its solar system, and the universe. Through these past missions, NASA has acquired a massive volume of data stored on an estimated 1 million reels of magnetic tape.

Because future missions are expected to produce additional volumes of data unparalleled in NASA's history, these missions will be forced to rely on the use of advanced data storage technologies. For example, one of NASA's future missions, EOS, is expected to provide over 2 terabytes[2] of data every day. Storing this daily supply of data would require about 10,000 reels of standard magnetic tape, which if stacked, would be nearly twice the height of the Washington Monument. In contrast, one of the advanced storage media now available—optical tape—could store the same amount of data on two reels, about 2 inches high.

[2]One terabyte of data is approximately 10^{12} bytes (or 1 trillion bytes). A byte is equivalent to 1 character in a text file, such as the letter "a."

NASA Initiatives to Meet Its Future Handling and Storage Requirements Will Rely on Advanced Technologies

Future space missions will last longer and carry an increasing number of sophisticated instruments producing more data. NASA estimates the volume of data it archives will rise over 5,600 percent from 1990 to 2000. Much of this increase is expected to begin with the 1998 launch of the first EOS platform, growing as the program becomes fully operational. EOS is NASA's main contribution to the international effort, which is studying the earth on a global scale in order to understand the processes that operate on the world's atmosphere, oceans, land surfaces, and other ecosystems. The program is significant in part because of the amount of data it will generate, producing about 70 percent of the data to be archived by the year 2000. See appendix II for more information on the growth of data from future missions.

Because its current ground processing and data storage systems will be unable to support future requirements, NASA is pursuing a number of major initiatives including EOSDIS, the Customer Data and Operations System (CDOS), advanced data storage technologies, and others.

EOSDIS, a key component of the EOS program, will direct the space platforms, provide the scientific community access to data from EOS, conduct advanced processing, and archive the data needed to support the global change research program. Officials in NASA's Office of Space Science and Applications could not provide a cost estimate of the EOSDIS portion of the EOS program. However, NASA's Administrator testified in May 1990 that about 60 percent of the EOS budget will be spent for EOSDIS, support to scientists, education programs, and related activities. The EOS program will grow from $191 million in fiscal year 1991 to an estimated $2.2 billion in fiscal year 2000, for a total estimated cost of $15.5 billion during the period. Total cost for the program is estimated to be $38.5 to $43 billion over its 15-year life.

CDOS will support EOS and other missions by providing initial ground processing, communications, data distribution, and serving as an archive for the initially processed data. Total costs for completion of CDOS and related improvements are estimated at $500 to $600 million.

Another initiative is encouraging the use of advanced storage technologies to handle the increasing data volumes expected by future missions. In the past, many missions relied on the older 9-track magnetic reel tape. In the future, however, missions plan to use newer storage methods, such as advanced optical and helical scan technologies, because they can store data more densely and are generally less costly than 9-track tape.

One of these, the 19 millimeter (mm) rotating helical scan recording technology, promises data storage users an increase in storage capacity that is 30 times the present levels of the 4mm and 8mm tapes. Although not yet commercially available, it has storage capacities comparable to most optical media, but with greater cost savings.

Even smaller-scale missions may benefit from advanced storage technologies. According to a NASA information systems official, every mission considers the use of advanced storage technologies, such as optical and helical scan, as part of its planning process. We reviewed project data management plans and other documents from selected future missions with high and low data volumes and found that all of the selected missions are considering, or plan to use advanced storage media. NASA is also funding other smaller initiatives to help meet the requirements of future high volume missions. These range from using/researching data compression techniques (ways of representing data with fewer bits) to acquiring optical disk equipment. NASA's projects and initiatives are discussed more in appendix III.

Future Challenges That Need to Be Prudently Managed

NASA faces significant data management and technology challenges in its quest to process and store the huge amounts of data expected during this decade. These challenges will involve

- being an early user of advanced mass storage systems that are not yet commercially proven, and thus may not work as expected;
- using technologies for which no media or format standards yet exist, and
- designing and developing EOSDIS, an information system of unprecedented size, complexity, and cost.

Unproven Storage System

Most of NASA's past missions stored data on standard magnetic tapes, which are relatively limited in their data holding capacity. Storing the large amounts of expected future data requires new and more efficient mass storage media. Several of the higher capacity media, however, have yet to be proven in the commercial market. One advanced storage medium is the 19mm helical scan tape, which, when placed in cartridges and used in a mass storage system, promises both a high storage capacity and the ability to quickly access data.

While the 19mm tape is commercially available and is used in the broadcast industry for video storage, the 19mm helical scan mass storage

system for digital data is not. This system still requires the development of new hardware and software components, including its robotic storage and retrieval mechanisms. The system is being developed by two manufacturers, and is expected to be on the market by 1992. However, some knowledgeable users at NASA and other agencies with large data bases question whether the technology will work as expected. A Goddard data systems official stated that questions still remain about how well the 19mm helical scan mass storage system will work, given prior unsatisfactory experience with other mass storage systems. For example, the official referred to another agency that tried to use a mass storage system. However, that system failed. According to a representative of the agency involved, the mass storage system failed because the small number of systems sold forced the manufacturer to stop supporting the system. He also stated that part of the problem was that the helical scan read/write components wore out frequently and were expensive to refurbish or replace.

Officials responsible for advanced technologies at a federal agency with one of the country's largest data storage systems, and a Jet Propulsion Laboratory engineer familiar with advanced storage systems, told us they were not aware of any successful advanced mass storage systems but knew of several that had failed.

Undefined Media and Format Standards

Two other challenges, which are important aspects of managing data, are the standards for (1) the physical media on which data are stored, and (2) the data's format. Standards are important in simplifying and reducing the cost of using data from various sources such as EOS. In the past, users had difficulties exchanging and comparing information from separate missions or from several branches of science because of a lack of data format standards.

Physical media standards have not been established for all commercially developed advanced data storage technologies. In particular, they do not exist for advanced technologies such as 19 mm helical scan tapes or optical tape. Additionally, format standards for all advanced technologies except compact disk-read only memory (CD-ROM) either do not exist, or if drafted have not been formally adopted by recognized national or international committees. The process of defining and adopting each standard can be rather lengthy, taking several years. NASA is involved with the accreditation of new standards, and provides information to missions about standards that exist or have been drafted.

A Complex Data Management System

A further formidable challenge will be pulling together diverse information from many scientific disciplines into a reliable system. EOSDIS, a key to the success of EOS, is to overcome a major past impediment in performing space and earth science research caused by the lack of information about what data were available, what their characteristics were, and how to obtain suitable versions of the data. EOSDIS will be a data-intensive system that will eventually manage more data than any other system in existence today. For example, the sheer volume of data—more than 1,000 times the amount of information in the Library of Congress—will be difficult to manage.

Historically, attempts to build information systems for large data volume applications failed because of their size and complexity. In this regard, NASA believes one of the largest challenges for the EOS program is EOSDIS. Additionally, the National Research Council has stated that EOSDIS will be a pioneering effort, far exceeding any existing civilian information system.[3] The Council also reported that the existing systems' performance in managing complex compilations of scientific data is not encouraging. For example, according to the Council, responses to data requests can be slow, and NASA data sets are known to be difficult to obtain.

Cost and Progress of Information Technology Initiatives May Not Be Visible to Congress

Key information technology initiatives like EOSDIS and CDOS must be visible to the Congress to permit effective oversight and informed decision making. NASA's current reporting mechanism—project status reports—are designed to keep congressional oversight committees informed about major NASA programs. However, because EOSDIS is part of a larger project and CDOS is considered an upgrade to an existing system, NASA's Comptroller's office does not plan to use project status reports to make Congress aware of these initiatives.

NASA has prepared project status reports on various projects since the mid-1970s. These reports include detailed information on project cost, schedule, and technical information. By mutual agreement between NASA and its four congressional oversight committees, a report is required for

[3]The U.S. Global Change Research Program: An Assessment of the FY 1991 Plans, National Academy of Science's National Research Council, National Academy Press, 1990.

(1) all new projects that meet NASA's definition of a project, and (2) are estimated to cost $200 million or more for research and development.[4]

The EOS program was given congressional new-start approval for fiscal year 1991, and a project status report will be due as of March 1991, after it reaches the $200 million threshold. However, according to a NASA comptroller official, the EOSDIS portion of EOS, which may cost billions, is not required to be separately disclosed in the EOS project status report. The CDOS program is estimated to cost $500 to $600 million to complete. According to a NASA comptroller official, however, new-start approval is not required because NASA does not consider CDOS a project but rather an upgrade of its current communication and data systems. Therefore, a project status report will not be prepared.

Opportunities Exist to Save Data Storage Costs

NASA should remain alert for opportunities to reduce potential costs. We found one such opportunity. Although NASA has not finalized the plans for EOSDIS, design documents show that NASA plans to archive potentially massive amounts of identical data at two locations. Although backup storage of data is prudent, this level of duplicate storage, covering a large portion of the estimated EOSDIS holdings, is not necessary and could raise acquisition costs as much as $35 million.

Specifically, after initially processed data is created at the CDOS facility in White Sands, New Mexico, it will be transmitted to the various EOSDIS data centers and other facilities throughout the country for further processing. After transmission, it will be archived at the CDOS facility as a backup copy.

The EOSDIS centers will then process the data received, creating several higher-level data products. However, after this processing, the original data received from CDOS will also be archived at the EOSDIS centers, reportedly to reconstruct the higher-level data products if they were ever lost or destroyed. This storage is unnecessary, however, because (1) most original data, according to an EOSDIS official, can be reconstructed directly from the higher-level data products, and (2) if for some reason the higher-level data products could not be used for reconstruction, the original data could be requested from the CDOS facility at White Sands.

[4] Project status reports on eligible projects are prepared biannually—as of March 15 and July 30—so they are available for use both at the beginning of the budget review process and at the end during the final framing of the spending bills. They are submitted during the first March following congressional new-start approval, or later when they reach the $200 million cost threshold.

If not eliminated, the additional costs to store the data in EOSDIS could be significant. At the three largest EOSDIS centers, original data will account for over 29 percent of all data to be stored within all centers for the initial 15-year period. We estimate that the additional cost to store this data would be approximately $35 million.[5]

After we raised this issue, statements and correspondence from EOSDIS officials indicated that they will eliminate the unnecessary data storage. However, at the time we completed our work in December 1990, the preliminary EOSDIS specifications had not been changed.

Conclusions

NASA has several initiatives underway to handle the processing and storage of future space and earth science data that will be generated at unprecedented rates and volumes. NASA believes that current initiatives in advanced storage media and in programs such as CDOS and EOSDIS, combined with other initiatives, will meet future demands to process and store the expected deluge of data.

CDOS and EOSDIS are information technology initiatives critical to NASA's ability to efficiently and economically process, analyze, and store the massive amounts of earth and space science data expected from future missions. However, NASA will face many formidable development challenges before all future processing and storage demands can be met. In addition, opportunities may exist to minimize developmental costs on these initiatives.

Further, despite their costs and risks, NASA's progress and problems with initiatives like EOSDIS and CDOS may not be readily visible to oversight committees. Given their importance, separate reporting on these initiatives appears prudent.

Recommendations

We recommend that the Administrator, NASA, review the current plans for EOSDIS and CDOS and ensure that unnecessary data holdings are eliminated, and that project funding be reduced accordingly. Further, to keep congressional oversight committees abreast of key information technology initiatives, we also recommend that the Administrator require NASA comptroller officials to either (1) prepare separate project status

[5]This estimate was based on information provided by manufacturers and covers initial acquisition costs for the data storage hardware and media.

reports on EOSDIS and CDOS, or (2) make their costs and progress clearly visible within an existing project status report.

gency Comments

In its written comments on a draft of this report, NASA neither agreed nor disagreed with our conclusions. NASA said it recognized the difficulty associated with the large projected data volumes and rates of future missions. NASA said it is facing this challenge squarely—planning to earmark a large portion of the EOS program costs for support of ground operations, data processing and storage, and scientific analyses. While recognizing the technological challenges highlighted in the report are real and involve risks, NASA expressed confidence for proceeding with its current approach, given that 7 years remain before the launch of the first EOS platform. Further, although NASA's written comments did not specifically address our recommendations, a subsequent discussion with a NASA headquarters official clarified that NASA plans to (1) change the EOSDIS specifications to eliminate the unnecessary data holdings, and (2) be vigilant in identifying other unnecessary requirements for EOSDIS and CDOS. NASA chose not to comment on our recommendation concerning separate disclosure to the Congress of such projects as EOSDIS and CDOS. NASA's comments are included as appendix IV.

As arranged with your office, unless you publicly announce the contents of this report earlier, we plan no further distribution until 30 days from the date of this letter. We will then give copies to other appropriate congressional committees, the Administrator of NASA, and other interested parties upon request.

This work was performed under the direction of Samuel W. Bowlin, Director for Defense and Security Information Systems, who can be reached at (202) 275-4649. Other major contributors are listed in appendix V.

Sincerely yours,

Ralph V. Carlone
Assistant Comptroller General

Contents

Table

Figures

Abbreviations

CD-ROM	compact disk-read only memory
CDOS	Customer Data and Operations System
CODMAC	Committee on Data Management and Computation
DD-2	data digital-2
EOS	Earth Observing System
EOSDIS	EOS Data Information System
GAO	General Accounting Office
GB	gigabyte
IMTEC	Information Management and Technology Division
mm	millimeter
NASA	National Aeronautics and Space Administration
NSSDC	National Space Science Data Center
TB	terabyte
WORM	write once read many

Objectives, Scope, and Methodology

The Subcommittee on Science, Technology, and Space, Senate Committee on Commerce, Science, and Transportation, asked us to identify (1) relevant advanced data storage technologies, including their strengths, weaknesses, and costs; and (2) NASA's initiatives for using these advanced technologies. As agreed, we addressed the first area in a previous report.[1] This report discusses NASA's initiatives. Additionally, it discusses several challenges NASA will face in the 1990s and beyond, the visibility of these important initiatives to the Congress, and examines whether opportunities exist for reducing development costs on the initiatives.

To identify NASA's initiatives for using advanced storage technologies, we

- conducted interviews with NASA project managers and headquarters' officials, and
- reviewed the NASA Research and Technology Objectives and Plans Summaries for fiscal years 1989 and 1990, and other agency documents for information regarding NASA's efforts to handle large data volumes from future missions and to encourage the use of advanced storage technologies.

To identify data management and storage challenges that NASA will face in handling large future data volumes in the 1990s and beyond we

- interviewed NASA headquarters, Goddard Space Flight Center, and Jet Propulsion Laboratory officials, as well as other federal agency officials familiar with mass storage systems;
- reviewed NASA studies and documents, National Research Council reports, and previous GAO reports; and
- attended the Second International American Institute of Aeronautics and Astronautics/NASA symposium on space information systems.

Storage media characteristics and other information on advanced storage media were obtained from our previous report. However, as part of our work we updated projections on future data volumes, based on information obtained from NASA's Office of Space Science and Applications. In addition, manufacturers and data storage consultants provided details on mass storage systems presently available and on the 19mm data digital-2 (DD-2) data storage technology currently under development. Manufacturer product specifications were used as a primary data

[1] Space Data: Information on Data Storage Technologies, (GAO/IMTEC-90-88FS, Sept. 12, 1990).

source. Information on technology performance was accepted as provided, and no testing of equipment was done to verify the accuracy of manufacturers' claims or other reports of performance.

Our audit work was performed in accordance with generally accepted government auditing standards, between July 1990 and December 1990 at various locations including NASA headquarters in Washington, D.C.; the Goddard Space Flight Center in Greenbelt, Maryland; and the Jet Propulsion Laboratory in Pasadena, California.

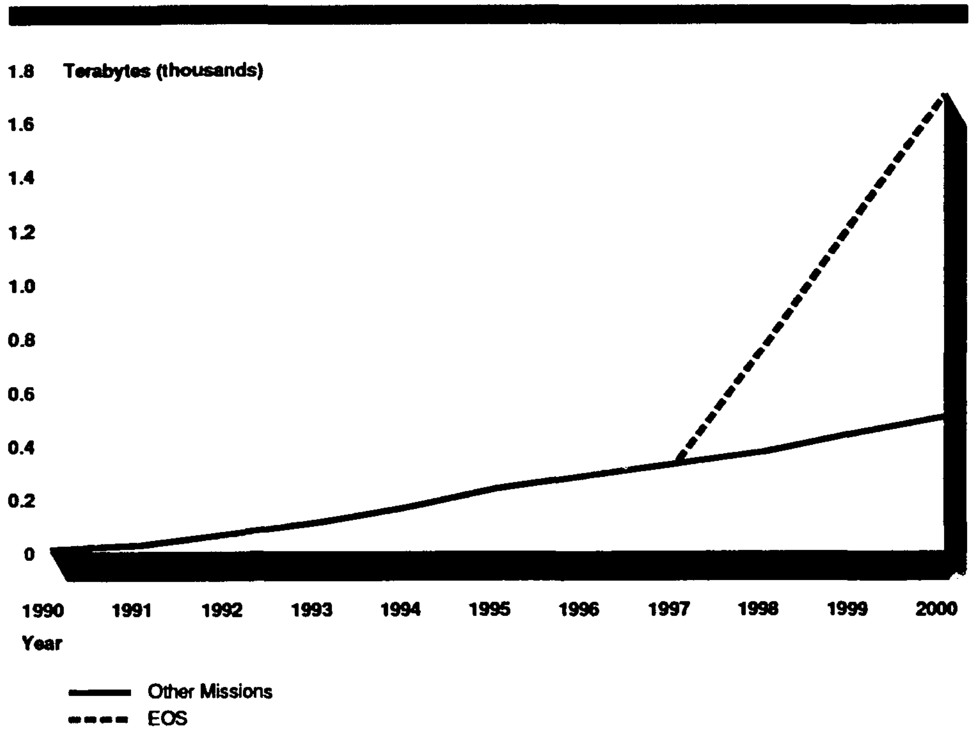
Greatly Increased Data From Future Missions

In the past, NASA space and earth science missions were typically single-purpose, data rates from instruments were generally low, and observing lifetimes were relatively short. Consequently, the amount of data produced was modest. In the future, however, NASA estimates that its missions will carry a greater number of complex instruments that will generate more data at sharply higher rates for much longer periods of time. The quantities of space and earth science data that NASA plans to generate, process, and store during the next decade will far exceed the quantities handled up to now. For example, the Magellan spacecraft, launched in May 1989, alone will generate more data than all previous planetary missions combined.

It is estimated that the annual volume of archive data will rise from about 8 terabytes of data in 1990 to about 460 terabytes by the year 2000—over a 5,600-percent jump.[1] This will result in the accumulated data volume climbing to an estimated 1,697 terabytes (as shown in figure II.1) with the 1998 launch of the first EOS platform during this period.

[1]Our previous report (GAO/IMTEC-90-88FS, Sept. 12, 1990) indicated a 6,500-percent increase between 1990 and 2000. This estimate, provided by NASA, was based on the EOS platforms A and B being launched in 1997 and 1999. Because NASA has revised the proposed launch dates to 1998 and 2001 respectively, the amount of data to be archived between 1990 and 2000 has dropped.

Figure II.1: NASA's Accumulated Archive Data Volume 1990 - 2000



Beyond the year 2000, the volume of data will be even greater as the EOS program becomes fully operational. EOSDIS—the archive for EOS data—is estimated to receive at least two terabytes of data per day on a continuous basis over the predicted 15-year life of the mission. This is expected to result in about 11,000 terabytes of data over the life of the mission, more than 1,000 times the text stored by the Library of Congress. Further, this figure could double or triple as data are reprocessed during analysis and if a data purging philosophy is not adopted.

Data Management and Storage Initiatives

The enormous increase in data volume from planned space and earth science missions in the 1990s will exceed NASA's current information system's ability to support future data handling and storage requirements. NASA has undertaken several major and other smaller-scale initiatives to meet the anticipated volume of future data.

Major Initiatives

The principal initiatives include EOSDIS, CDOS, an upgrade of a computer center at the Goddard Space Flight Center, and the encouragement of advanced data storage technologies.

EOSDIS

EOSDIS is a major NASA initiative to handle future data volumes, and a key component of EOS. After initial processing, EOSDIS will receive data from two EOS orbiting platforms each carrying a proposed complement of 14 science instruments, with additional data coming from other platforms and satellites. The first platform, EOS-A, is scheduled to be launched in 1998, while a second platform, EOS-B, is scheduled to be launched in 2001. The EOS synthetic aperture radar, once part of EOS-B, is now scheduled for a 1999 launch on its own platform. The European and Japanese space agencies, as part of the international effort, will launch polar orbiting platforms in the late 1990s. Other satellites, known as earth probes, will provide preliminary measurements on specific earth processes and are scheduled to be launched prior to the EOS platforms.

EOSDIS is planned to provide the scientific community access to data from EOS and other platforms, for use in global change research. The information system is intended to furnish the capability to control the platforms, distribute data to users, perform advanced level processing,[1] and archive the data.

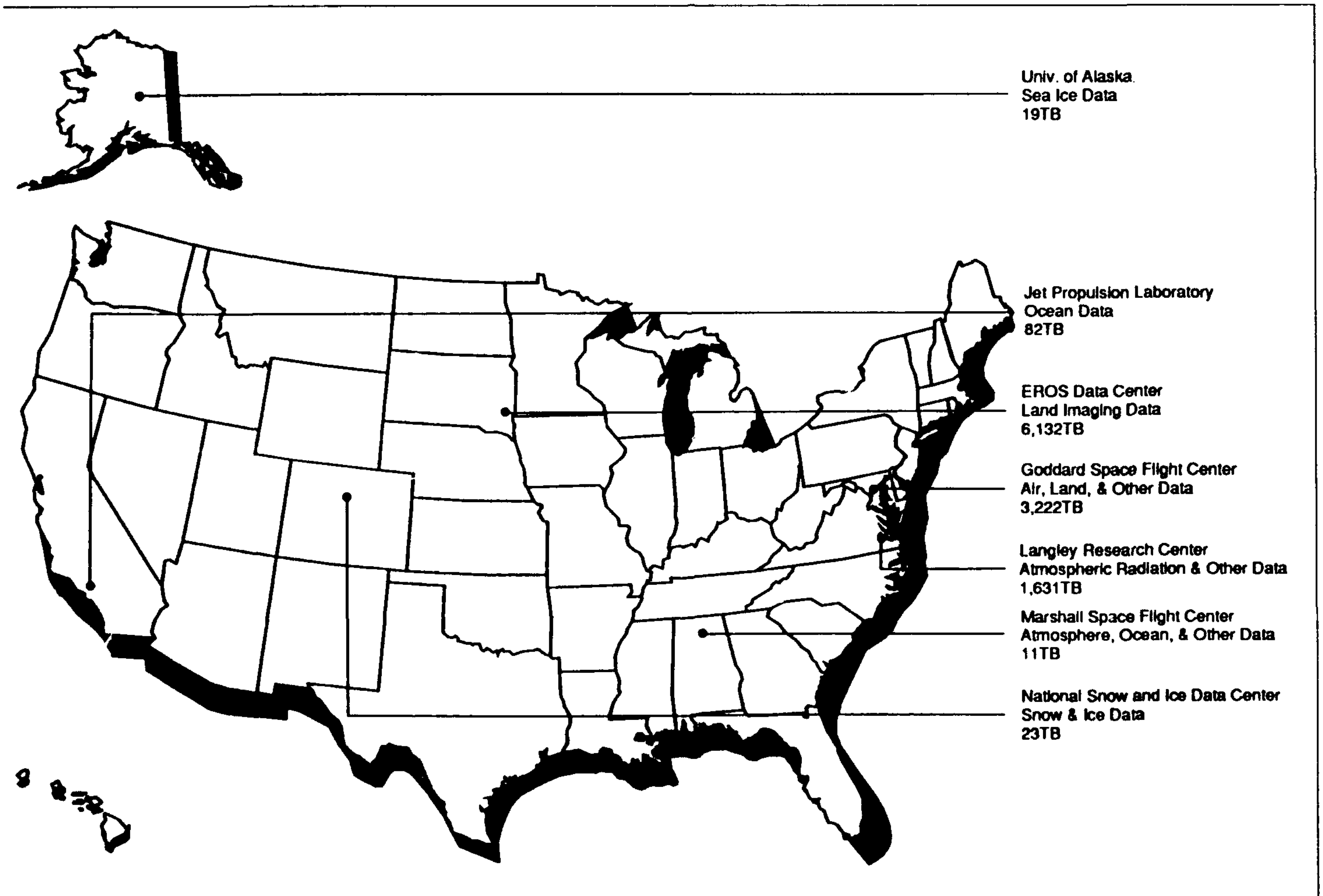
EOSDIS will handle EOS data from end-to-end (spacecraft to researchers to archive) using a combination of existing and planned NASA resources, and other capabilities and facilities developed specifically for EOS. For example, data will be transmitted from space and captured using NASA's current tracking and data relay satellites, and the White Sands ground terminal in New Mexico. Dedicated EOS services and facilities will be known as the EOSDIS "core system." As part of the core system, EOS data will be stored and managed at seven distributed active archive centers located around the country. The centers were chosen based on their expertise in particular subjects. For example, sea ice data will be housed

[1] Advanced level processing converts the raw scientific data to a usable form for data analysis.

at the University of Alaska. Each center will carry out advanced processing, distribution, and storage functions.

Data will not be distributed equally among the centers, so each center will have different storage requirements. Three sites, Goddard Space Flight Center in Greenbelt, Maryland; Earth Resources Observation Systems Data Center in Sioux Falls, South Dakota; and Langley Research Center, Hampton, Virginia, will maintain far greater volumes of data. Figure III.1 shows the location of each center, and the volume and type of data expected.

igure III.1: Location, Type, and Estimated Volume of Data to Be Stored at the Seven Earth Observing System Distributed Active rchive Centers



Note: Estimated volumes of data are expressed in terabytes (TB).

Because of the high data volume at these centers, the demands on the mass storage system selected for those centers will be severe. Only one mass storage system under development claims the capability needed for these large centers. For those with lower volume, other existing mass storage systems may be suitable. Table III.1 compares the storage capacity of several mass storage systems.

Table III.1: Comparison of Mass Storage Systems

Mass Storage System Manufacturer	Medium Type	Number Media Held	Total System Capacity
Optical Disk System - 6800 Automated Disk Library (Eastman Kodak)	WORM 14" disk (6.8GB capacity)	150 disks	1TB+
M960 Storage Management System with M1000 Storage Modules (Masstor)	1" helical scan magnetic tape cassette (31.25GB capacity)	256 cassettes	8TB+
Rotary Storage System 600 (Metrum)	13mm helical scan magnetic tape cassette (10GB capacity)	600 cassettes	6TB
Automatic Cassette Library System (Ampex)	19mm DD-2 helical scan magnetic tape cassette (25GB capacity)[a]	256 cassettes	6.4TB+
Automated Cartridge System - 4400 (StorageTek)	18-track half inch tape cartridge (3480 compatible; .2GB capacity)	96,000 cartridges	19TB+
EMASS³00 (E-Systems)	19mm DD-2 helical scan magnetic tape cassette (75GB capacity)[b]	99,176 cassettes	7,000TB

[a]The 19mm DD-2 has three capacity sizes: small, medium, and large. The Ampex sizes are small, 25GB; medium, 75GB; and large, 165GB. The Ampex system will only use the small cassettes.

[b]As mentioned in note a from Table III.1, the 19mm DD-2 tape has three tape sizes. The E-Systems capacity sizes are also: small, 25GB; medium, 75GB; and large 165GB. All three tape sizes can be used in the EMASS³00 system, but the system can automatically mount the small and medium cassettes only. The large cassettes must be inserted manually. The system is described here based upon use of the medium-sized cassettes—i.e. the largest capacity cassette the system will mount automatically.

The EOS information system will be massive, complex, and expensive. Data volumes are expected to exceed two terabytes per day and could go as high as four terabytes per day over the 15-year life of the mission. Further, a data management system as large and complex as EOSDIS has never been built. It will require a powerful data base management system in order to support even the most basic needs of the project, such as finding and accessing data requested by a user. NASA plans to spend a significant amount of money for this effort. The estimated cost for the EOS over its lifetime is between $38.5 and $43 billion, according to calculations made from testimony by the NASA Administrator in May 1990. Of that amount, 60 percent is for EOSDIS, support of scientists, education

programs and related activities. Separate figures for EOSDIS covering the life of the program were not available.

EOS was approved as an earth science mission new-start for the fiscal year 1991 budget submitted to Congress. Preliminary studies were funded under NASA's research and development budget, according to an EOS project official. NASA is currently preparing a "request for proposals" for the development of the EOSDIS core system, with a contractor expected to be selected by mid-February 1992. The system is scheduled for operation by mid-1998, prior to the first EOS platform launch.

CDOS

A second NASA initiative to handle future data rates and volumes is CDOS. CDOS will upgrade data capture, initial ground processing, communications, and data distribution. It will also serve as an archive for the initially processed data. CDOS is located at two facilities, with 70 percent of the processing and storage hardware located at the White Sands Complex. The management, operations, and sustaining engineering will be located at Goddard Space Flight Center.

CDOS is designed to improve NASA's current decentralized capabilities to process and store data. The current system is not capable of handling future data rates and volumes, whereas CDOS is intended to meet these future requirements. Performance requirements for CDOS are to handle 925 megabits per second of data from space, and to archive 6,000 terabytes of data over 10 years.

The current facilities provide service by customizing or developing unique processing and distribution for each particular mission. Instead of building a custom system for individual missions, CDOS uses centralized and common hardware and software for all supported missions. NASA expects major cost savings using CDOS.

NASA estimates the cost of CDOS, and related improvements to the communications system, at $500 to $600 million through completion. NASA is currently in the process of soliciting proposals for the design and implementation phase of the CDOS project. A contract for the work is expected to be awarded by mid-1991, and the CDOS system is scheduled for completion by mid-1996.

Computer Center Upgrade

A third initiative undertaken by NASA to handle the storage and processing needs for future missions is the upgrading of NASA's Center

for Computational Sciences mass storage capabilities. The computer center, located at the Goddard Space Flight Center, provides support for the analysis and modeling of data for scientists from all NASA science disciplines. The need for computer services are expected to grow as scientists develop environmental models for EOS data, using data from past missions to test the models.

Currently, the center is procuring a mass data storage and delivery system to meet these expanding archive retrieval and data analysis needs. A contract is expected to be awarded by late spring or early summer 1991, with the upgrade to be delivered 90 days after award. The initial upgrade is from a 0.5 terabyte capacity system up to 7.2 terabytes, with the ability to grow to 229 terabytes. The budgeted cost for the 7.2 terabyte system is procurement sensitive and cannot be released until the contract is awarded, according to the head of the computer center.

Advanced Data Storage Technologies

We consider NASA's encouragement of missions to use advanced storage technologies as a fourth initiative for handling the increasing data volumes expected in the future. Many past missions used the older 9-track magnetic reel storage technology. Future missions, however, plan to use advanced optical and helical scan technologies for some storage needs. These advanced technologies store more data on fewer media and are generally less costly than 9-track tape.[4]

NASA is encouraging the use of advanced storage technologies by offering seminars and issuing reports on advanced storage media as well as requiring each mission to outline data storage and handling procedures in a project data management plan.[5] The space agency is also establishing a clearinghouse function at NASA's National Space Science Data Center (NSSDC) to assist missions in selecting appropriate advanced media based on the mission's storage needs and data volume.

Future missions with both high and low data volumes plan to use advanced optical and helical scan technologies because of capacity and

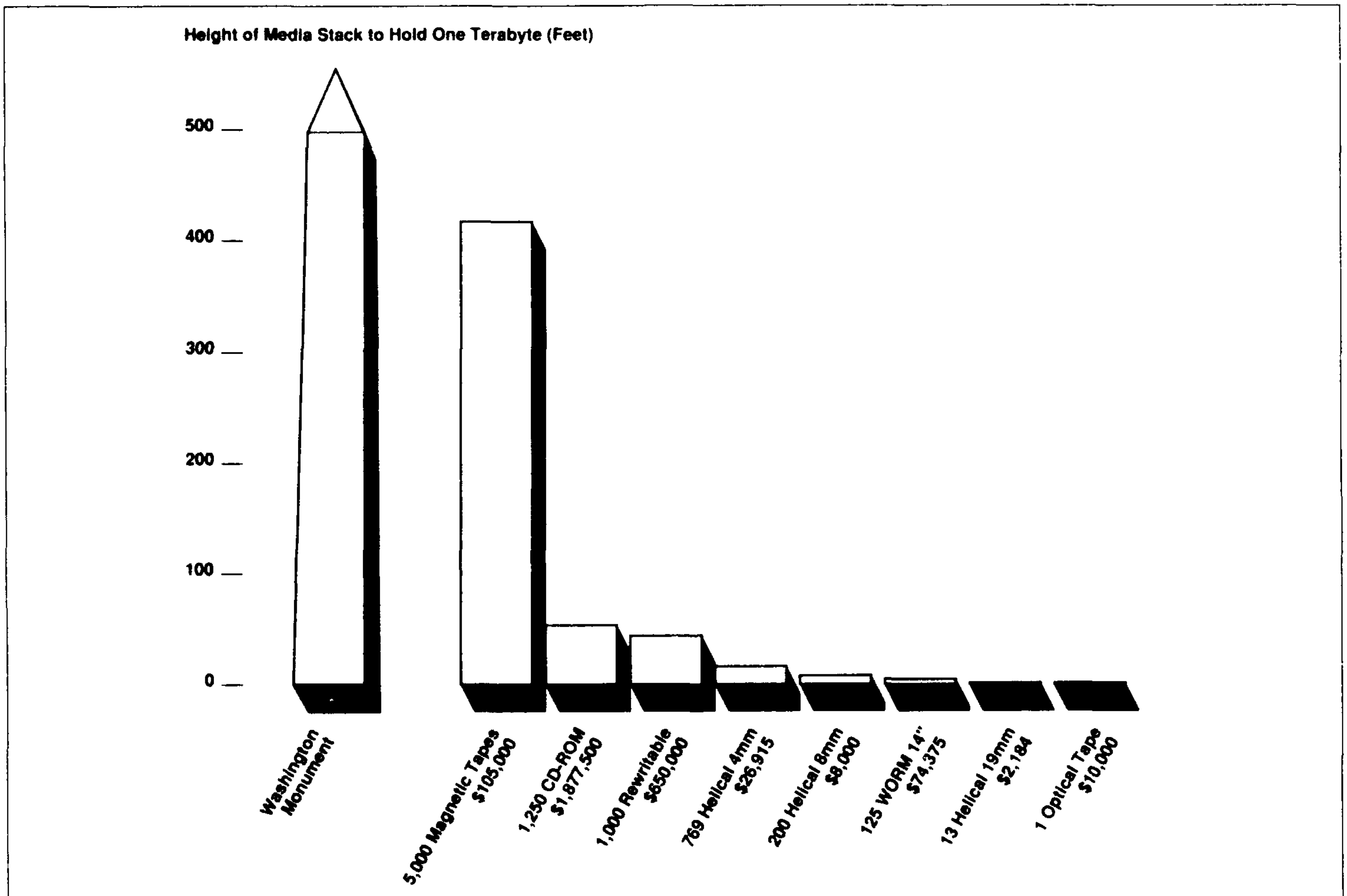
[4]For further information on the characteristics and costs of advanced storage technologies see GAO's report Space Data: Information on Data Storage Technologies (GAO/IMTEC-90-88FS, Sept. 12, 1990).

[5]This plan describes mission goals and expected data volumes, as well as data storage and handling procedures. It must be approved 6 months prior to mission launch in accordance with NASA Management Instruction 8030.3A. The instruction is currently being revised, but is expected to retain the requirement for the plan.

cost advantages. Optical or laser recorded storage media provide longer storage life, generally lower cost-per-megabyte, and less frequency of errors than 9-track tape. Laser recorded media includes compact disk-read only memory (CD-ROM) disks, write once read many (WORM) disks, rewritable optical disks, and optical tape.

Advanced storage systems using the rotating helical scan recording technology have storage improvements over 9-track tape comparable to most optical media, but generally with greater cost savings. Current helical scan products include 4mm and 8mm magnetic tape cassettes. However, another helical scan recorder system using the 19mm DD-2 magnetic tape is expected to be commercially available for data storage by 1992. The system employs the same 19mm tape presently used in video recorders by the broadcast industry. It promises an increase in both storage capacity and data rate that is 30 times the current levels of the 4mm and 8mm tapes. Figure III.2 compares the capacity and cost of optical and helical scan technologies.

Figure III.2: Number of Media and Cost to Store One Terabyte of Data on Magnetic Reel and Advanced Media



Source: GAO adaptation of an E-Systems illustration

Note: The CD-ROM cost reflects the cost of mastering ($1,500 per disk) and of one copy ($2.00). Subsequent copies cost $2.00 each.

According to a NASA information systems official, every mission considers the use of advanced storage technologies, such as optical and helical scan, as part of its planning process. We reviewed project data management plans and other documents from selected future missions with high and low data volumes and found that all of the selected missions are considering, or plan to use, advanced storage media in some capacity. For example, the EOS mission, with more data to be archived than all other future NASA missions, is considering the use of 19mm DD-2 tape in a mass storage system. The Magellan mission to Venus, the

highest data generator for solar system missions, anticipates using WORM disks for mass data storage and retrieval. And the Cosmic Background Explorer, a relatively low data generator, plans to use helical 8mm tape cartridges as an intermediate or temporary storage medium.

ther Initiatives

In addition to the above initiatives, NASA is also funding other smaller-scale initiatives to help meet the demands of a high volume data future. These range from the development of spaceflight data compression techniques with applications for ground-based storage, to the purchase and distribution of equipment for creating and using optical disks. According to a NASA information systems official, other NASA projects may have elements addressing data storage technologies, however he could not readily identify them.

ata Compression

The Jet Propulsion Laboratory is developing compression technology for storing data on-board the spacecraft at a 2 to 1 compression ratio. The technology could eventually be applied to ground storage of data, with the practical result of requiring less storage space.

Goddard Space Flight Center is working on two data compression approaches, one that compresses data at a 2 to 1 ratio, and the other at a 20 to 1 ratio with some loss of data in the process. The project will test the effects of both types of compression on the scientific analysis of data.

rasable Optical Disk ecorder

Langley Research Center is constructing a high capacity, high transfer rate, erasable optical disk recorder for on-board storage and retrieval of data. The technology will be applicable to the development of ground-based optical disk drives.

D-ROM Premastering ervices and Facilities

NSSDC offers equipment, advice, and space for the preparation of data to produce a CD-ROM master disk, from which many copies can be cheaply made.

VORM Optical Disk quipment

NSSDC obtained a drive system that will allow the movement of data from tape images to optical disk, and back to tape, allowing data to be stored on optical disk and to be distributed to users on tape if optical equipment is not available to them.

NSSDC purchased 40 WORM optical disk drives and distributed them to researchers in the space community. Principal investigators with the equipment then submit data for archiving to NSSDC on optical disk.

Migrating Older Data Sets to Higher Density Media

NASA has been migrating older data sets from several missions launched during the 1970s to smaller, higher density media. For example, most recent plans are for Hawkeye plasma physics data and Skylab X-rays of the sun to be transferred to optical disk, and Viking and Voyager planetary data to CD-ROM.

Standards Office

In 1988, NSSDC's Standards Office was established to promote the use of standard data formats for the archiving, exchange and storage of space and earth sciences data, and the development of new standards. Standards are important in simplifying and reducing the cost of using data from various sources. An on-line standards information service began operating at the end of 1990. Users can access the data base using a local network, or request a printout of information. User support is available for assistance in applying a standard for image data, or for using the international standard format data unit.

CD-ROM Distribution and Archive Laboratory

The Jet Propulsion Laboratory is developing a data distribution and archive lab to provide assistance to scientists in distributing new or existing data sets on CD-ROM. Other lab tasks are to evaluate storage media and hardware costs and benefits for distributing data and storing it for long periods, and to participate in the development of standards for storage technology.

omments From the National Aeronautics and pace Administration

NASA

National Aeronautics and
Space Administration

Washington, D.C
20546

Office of the Administrator

FEB 28 1991

Mr. Ralph V. Carlone
Assistant Comptroller General
Information Management and
Technology Division
United States General Accounting Office
Washington, DC 20548

Dear Mr. Carlone:

This is the National Aeronautics and Space Administration's (NASA) response to the General Accounting Office (GAO) Draft Report/IMTEC-91-24, entitled "Space Data: NASA's Future Data Volumes Create Formidable Challenges," dated January 28, 1991.

NASA appreciates the opportunity to comment on the draft report, and we also appreciate the efforts of your staff to solicit and consider our views through the process of the review.

NASA recognizes the difficulty associated with the large projected data volumes and rates for future science missions, and is treating this challenge seriously in mission planning for all our science disciplines. In the EOS program, for example, 60 percent of total funds will support ground operations, data and most importantly science. This represents a stronger commitment by NASA to address the data management challenge than has been made for any other program in the Agency's history.

The technological challenges outlined in the draft report are real and not without risk. Data storage requirements are receiving special attention. As listed in Appendix III of the draft report, several devices having appropriate storage capacities are in various stages of development and marketing. Meanwhile, the market for this class of mass data storage is growing rapidly, and over seven years remain before launch of the first EOS platform. Our assessment of the current situation provides confidence for proceeding with our implementation approach.

2

The draft report indicates that the storage requirements may be unnecessarily redundant for Level 0 data (i.e., original data records) from EOS instruments. The data collected for the instruments is critical to the overall success of the EOS mission and must be protected as such. Part of this protection is retention of backup copies to recover against loss and to assure integrity of datasets and recordings. This backup function was originally planned as a responsibility of the Distributed Active Archive Centers of the EOS Data and Information System (EOSDIS). However, a permanent archive of all Level 0 data is now being incorporated in the functional design of the Customer Data and Operations System (CDOS). Therefore, the EOSDIS will not store two copies of Level 0 data but will rely upon CDOS for the backup.

We agree that a vigil must be maintained for ways to minimize costs, and believe we are being careful in doing so. As we proceed with the implementation of our planned systems, we will continue to have opportunities to monitor the design and costs carefully.

The interest of the Congress and GAO in helping to highlight the challenges in meeting the needs for easily accessible and high quality science data products is greatly appreciated. We will continue to work diligently to provide these data products in a highly responsible and economical manner.

Sincerely,

for John E. O'Brien
Assistant Deputy Administrator

Major Contributors to This Report

Information Management and Technology Division, Washington, D.C.

Ronald W. Beers, Assistant Director

Los Angeles Regional Office

Allan Roberts, Assistant Director
George Vindigni, Evaluator-in-Charge
Monica Kelly, Staff Evaluator
Shawnalynn R. Smith, Staff Evaluator

[illegible]

[illegible] copies of each report are free. Additional copies are $[illegible] each.

There is a 25% discount on orders for 100 or more copies mailed to a single address.

Orders must be prepaid by cash or by check or money order made out to the Superintendent of Documents.